COMMUNICATIONS IN NUMERICAL METHODS IN ENGINEERING
Commun. Numer. Meth. Engng 2006; **00**:1–6 *Prepared using* cnmauth.cls *[Version: 2002/09/18 v1.02]*

A locally conservative least-squares method for Darcy flows

P. B. Bochev[1,*] and M. D. Gunzburger[2,‡]

[1] *Computational Mathematics and Algorithms, Sandia National Laboratories[†], P.O. Box 5800, MS 1110, Albuquerque NM 87185-1110*
[2] *School of Computational Science, Florida State University, Tallahassee FL 32306-4120*

SUMMARY

Least-squares finite element methods for Darcy flow offer several advantages relative to the mixed Galerkin method: the avoidance of stability conditions between finite element spaces, the efficiency of solving symmetric and positive definite systems, and the convenience of using standard, continuous nodal elements for all variables. However, conventional C^0 implementations conserve mass only approximately and for this reason they have found limited acceptance in applications where locally conservative velocity fields are of primary interest. In this paper we show that a properly formulated *compatible* least-squares method offers the same level of local conservation as a mixed method. The price paid for gaining favorable conservation properties is that one has to give up what is arguably the least important advantage attributed to least-squares finite element methods: one can no longer use continuous nodal elements for all variables. As an added benefit, compatible least-squares methods inherit the best computational properties of both Galerkin and mixed Galerkin methods and, in some cases, yield identical results, while offering the advantages of not having to deal with stability conditions and yielding positive definite discrete problems. Numerical results that illustrate our findings are provided.

KEY WORDS: Least-squares, finite element methods, mixed methods, Darcy flow, local conservation

1. Introduction

This paper considers least-squares finite element methods for the elliptic boundary value problem

$$\begin{cases} \nabla \cdot \mathbf{u} + \gamma\phi = f & \text{and} \quad \mathbf{u} + \mathbb{A}\nabla\phi = \mathbf{0} \quad \text{in } \Omega \\ \phi = 0 \quad \text{on } \Gamma_D & \text{and} \quad \mathbf{u}\cdot\mathbf{n} = 0 \quad \text{on } \Gamma_N\,. \end{cases} \tag{1}$$

In (1), Ω denotes a bounded open region in $\mathbb{R}^n$, $n = 2, 3$, with Lipschitz continuous boundary Γ consisting of two disjoint pieces Γ_D and Γ_N; $\mathbb{A}$ is a rank two symmetric, positive definite

*Correspondence to: Computational Mathematics and Algorithms, Sandia National Laboratories, P.O. Box 5800, MS 1110, Albuquerque NM 87185-1110
[†]Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC-94AL85000.
[‡]Supported in part by CSRI, Sandia National Laboratories, under contract 18407.

Received 31 March 2006
Revised 31 March 2006

tensor; $\gamma \in L^\infty(\Omega)$ and $f \in L^2(\Omega)$ are two given scalar functions[†]. We assume that there exists a constant $\alpha > 0$ such that for every $\mathbf{x} \in \Omega$,

$$\frac{1}{\alpha}\boldsymbol{\xi}^T\boldsymbol{\xi} \le \boldsymbol{\xi}^T\mathbb{A}(\mathbf{x})\boldsymbol{\xi} \le \alpha\boldsymbol{\xi}^T\boldsymbol{\xi}\,. \tag{2}$$

The properties of $\mathbb{A}$ ensure the existence of the square root $\mathbb{A}^{1/2}$ and its inverse $\mathbb{A}^{-1/2}$. We will treat two cases: either $\gamma(\mathbf{x}) \ge \gamma_0 > 0$ for $\mathbf{x} \in \Omega$ and $\|\gamma\|_{L^\infty(\Omega)} \le C$ for some constants $C > 0$ and $\gamma_0 > 0$, or $\gamma \equiv 0$. Two important cases for which $\gamma \ge \gamma_0 > 0$ are diffusion-reaction and parabolic problems. In the latter case, γ can be identified with $1/\Delta t$, where Δt is the time step used to effect the temporal discretization.

The boundary value problem (1) arises in applications such as porous media flow and semiconductor device modeling, where the velocity $\mathbf{u}$ is more important than the pressure ϕ. In such cases numerical methods that compute accurate, locally conservative velocity approximations are favored. Mixed Galerkin methods are one such approach. However, mixed methods yield less accurate pressure approximations and lead to saddle-point systems that are more difficult to solve[‡]. They also require spaces that are subject to an inf-sup condition and cannot be chosen independently of each other [5]. In particular, standard, equal-order C^0 finite elements form an unstable pair.

In least-squares finite element methods [1] the saddle-point variational problem is replaced by an unconstrained minimization of a quadratic least-squares functional defined by summing up the residuals in (1). Least-squares principles lead to symmetric and positive definite systems and remain stable for any conforming approximations of ϕ and $\mathbf{u}$, including equal order C^0 elements. The latter has been long considered to be one of the principal advantages and reasons to use least-squares methods. Unfortunately, when implemented in this way, least-squares methods for (1) are only approximately conservative and the L^2 velocity error is suboptimal [2, 11, 9]. Two different approaches were proposed to remedy the suboptimal velocity convergence. One was to augment (1) by a redundant curl equation [6, 7, 8, 12] so as to obtain an H^1 norm-equivalent functional. A second approach, [9], relied on the grid decomposition property of the uniform criss-cross grid to establish optimal L^2 velocity convergence without the redundant curl equation.

Unfortunately, least-squares methods based on the augmented first-order system do not approximate well solutions of (1) that are strictly in $H(\Omega, \mathrm{div}\,)$ and uniform criss-cross grids offer limited meshing capability for general domains. Because of these limitations least-squares methods in [6, 7, 8, 12] and [9] had trouble competing with mixed Galerkin methods for (1).

In this paper we show that a properly formulated *compatible* least-squares method can deliver optimal convergence rates in *all* variables. The term *compatible* underscores the fact that in

[†]Throughout this paper we use standard space notations. For $p > 0$, $H^p(\Omega)$ denotes the Sobolev space of order p with norm and inner product denoted by $\|\cdot\|_p$ and $(\cdot,\cdot)_p$, respectively. When $p = 0$, we use the standard notation $L^2(\Omega)$. The symbol $|\cdot|_k$, $0 \le k \le p$, denotes the kth seminorm on $H^p(\Omega)$ and $H^1_D(\Omega)$ is the subspace of $H^1(\Omega)$ consisting of all functions that vanish on Γ_D. Vector-valued functions and vector analogues of the Sobolev spaces are denoted by lower and upper case bold-face font, respectively, e.g., $\mathbf{u}$, $\mathbf{H}^1(\Omega)$, $\mathbf{L}^2(\Omega)$, etc. The set $H(\Omega, \mathrm{div}\,) = \{\mathbf{u} \in \mathbf{L}^2(\Omega) \mid \nabla\cdot\mathbf{u} \in L^2(\Omega)\}$, and its subset $H_N(\Omega, \mathrm{div}\,) = \{\mathbf{v} \in H(\Omega, \mathrm{div}\,) \mid \mathbf{v}\cdot\mathbf{n} = 0 \quad \text{on } \Gamma_N\}$ are Hilbert spaces when equipped with the graph norm $\|\mathbf{u}\|_{H(\Omega,\mathrm{div}\,)} = (\|\mathbf{u}\|_0^2 + \|\nabla\cdot\mathbf{u}\|_0^2)^{1/2}$.

[‡]Hybridization is a technique that allows the reduction of a saddle-point system to a smaller symmetric and positive definite equation [5]. However, hybridization procedures depend heavily on the spaces used in the mixed method and are developed on a case by case basis.

this method each variable is approximated according to the classical variational principle it was inherited from. The choice of approximation spaces is the single most important difference between compatible and conventional C^0 least-squares methods which gives rise to fundamental differences in their properties. Most notably, in compatible least-squares method each variable is not only approximated to the same accuracy as in the mixed-Galerkin and Ritz-Galerkin finite element methods, but under some conditions the solutions obtained by the different methods are identical. Therefore, when these conditions are fulfilled, the compatible least-squares method computes exactly the same locally conservative velocity approximation as the mixed method but without having to solve a saddle-point system. Remarkably, when these conditions are not all fulfilled, compatible least-squares principles remain meaningful and a simple post-processing procedure can be used to recover local conservation.

Besides the fact the compatible least-squares methods can deliver locally conservative velocity fields, another difference with C^0 methods is that optimal L^2 convergence of the velocity does not require the addition of a redundant curl equation. There are also important differences between compatible least-squares and mixed finite element methods. Indeed, compatible least-squares finite element methods do not require any stability conditions between the finite elements spaces so that they can be chosen completely independently from each other. In principle, the spaces do not even have to be defined with respect to the same mesh, which may prove useful in some situations. Of course, as usual, compatible least-squares finite element methods yield symmetric, positive definite algebraic systems that are easier to solve than the symmetric, indefinite systems obtained from mixed methods.

We have organized this paper as follows. In Section 2 we present the compatible least-squares method for (1) and prove that if $\gamma_0 > 0$ this method computes exactly the same pressure and velocity as the Ritz-Galerkin method and the mixed Galerkin method. The case $\gamma = 0$ is considered in Section 3 where we develop a simple post-processing procedure that recovers the local conservation property of the velocity approximation. Section 4 presents a computational study of the compatible least-squares method that highlights its valuable computational properties. This section also compares compatible and standard C^0 least-squares implementations when (1) has rough solutions.

2. A compatible least-squares finite element method

We will formulate the compatible least-squares method for (1) under the assumption that $\gamma_0 > 0$. The main idea is to use pressure and velocity spaces that are inherited from the Ritz-Galerkin and the mixed Galerkin methods, respectively. Assume that $\mathcal{T}_h$ is a regular partition of Ω into finite elements K and let $\mathbf{G}^h$ denote a finite element subspace of $H^1_D(\Omega)$. A Ritz-Galerkin method for (1) is given by: seek $\phi^h \in \mathbf{G}^h$ such that

$$\int_\Omega \nabla\phi^h \cdot \mathbb{A}\nabla\widehat{\phi}^h \, d\Omega + \int_\Omega \gamma\phi^h\widehat{\phi}^h \, d\Omega = \int_\Omega f\widehat{\phi}^h \, d\Omega \quad \forall\widehat{\phi}^h \in \mathbf{G}^h \,. \tag{3}$$

In the Ritz-Galerkin method the velocity approximation $\mathbf{u}^h$ is given by $\nabla\phi^h$, is less accurate than ϕ^h, and is not locally conservative.

Let now $\mathbf{Q}^h \subset L^2(\Omega)$ and $\mathbf{D}^h \subset H_N(\Omega, \mathrm{div}\,)$ denote a pair of finite element spaces such that $\mathbf{Q}^h = \nabla\cdot(\mathbf{D}^h)$ and the inf-sup condition for (1) is satisfied; see [4, 5]. The mixed Galerkin

method for (1) is given by: seek $(\mathbf{u}^h, \phi^h) \in \mathbf{D}^h \times \mathbf{Q}^h$ such that

$$\left\{ \begin{array}{rcll} \displaystyle\int_\Omega \mathbb{A}^{-1}\mathbf{u}^h \cdot \widehat{\mathbf{u}}^h \, d\Omega - \int_\Omega \phi^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega & = & 0 & \forall\, \widehat{\mathbf{u}}^h \in \mathbf{D}^h \\ \displaystyle -\int_\Omega \widehat{\phi}^h \nabla \cdot \mathbf{u}^h \, d\Omega - \int_\Omega \gamma \phi^h \widehat{\phi}^h \, d\Omega & = & \displaystyle -\int_\Omega f\widehat{\phi}^h \, d\Omega & \forall\, \widehat{\phi}^h \in \mathbf{Q}^h \, . \end{array} \right. \tag{4}$$

The approximation properties of the mixed method (4) are dual to the Ritz-Galerkin method (3). In (4) the pressure is approximated by discontinuous functions and is less accurate than the C^0 approximation in (3). On the other hand, $\mathbf{u}^h$ in (4) is locally conservative and is in general more accurate than the field obtained from (3).

The compatible least-squares method combines the best computational properties of (3) and (4) by using the same finite element space $\mathbf{G}^h$ as in (3) to approximate ϕ^h and the same finite element space $\mathbf{D}^h$ as in (4) to approximate $\mathbf{u}^h$. The spaces $\mathbf{G}^h$ and $\mathbf{D}^h$ are not subject to a joint compatibility condition and so we are free to choose these spaces completely independently from each other. In particular, $\mathbf{G}^h$ and $\mathbf{D}^h$ can be defined on different conforming finite element partitions of Ω.

The compatible least-squares principle for (1) is given by the minimization problem

$$\min_{\phi^h \in \mathbf{G}^h, \mathbf{u}^h \in \mathbf{D}^h} J(\phi^h, \mathbf{u}^h; f) \, , \tag{5}$$

where

$$J(\phi^h, \mathbf{u}^h; f) = \frac{1}{2} \left(\|\mathbb{A}^{-1/2}(\mathbf{u}^h + \mathbb{A}\nabla\phi^h)\|_0^2 + \|\gamma^{-1/2}(\nabla \cdot \mathbf{u}^h + \gamma\phi^h - f)\|_0^2 \right) . \tag{6}$$

Theorem 1. *The problem* (5) *has a unique minimizer* $(\phi^h, \mathbf{u}^h) \in \mathbf{G}^h \times \mathbf{D}^h$. *The scalar* ϕ^h *is the unique solution of the Ritz-Galerkin method* (3) *and there exists a scalar* $\lambda^h \in \mathbf{Q}^h$ *such that the pair* $(\mathbf{u}^h, \lambda^h)$ *is a solution of the mixed-Galerkin problem* (4). *The velocity field* $\mathbf{u}^h$ *is locally conservative in the sense that*

$$\Pi_{\mathbf{Q}^h}(\nabla \cdot \mathbf{u}^h + \gamma\lambda^h - f) = 0 \, , \tag{7}$$

where $\Pi_{\mathbf{Q}^h}$ *is the* L^2 *projection onto* $\mathbf{Q}^h$.

Proof. Using standard calculus of variations techniques, we see that minimizers of (5) solve the variational equation: seek $(\phi^h, \mathbf{u}^h) \in \mathbf{G}^h \times \mathbf{D}^h$ such that

$$Q(\{\phi^h, \mathbf{u}^h\}, \{\widehat{\phi}^h, \widehat{\mathbf{u}}^h\}) = P(\{\widehat{\phi}^h, \widehat{\mathbf{u}}^h\}) \quad \forall \{\widehat{\phi}^h, \widehat{\mathbf{u}}^h\} \in \mathbf{G}^h \times \mathbf{D}^h \, , \tag{8}$$

where

$$\begin{aligned} &Q(\{\phi^h, \mathbf{u}^h\}, \{\widehat{\phi}^h, \widehat{\mathbf{u}}^h\}) = \\ &\quad \int_\Omega (\mathbf{u}^h + \mathbb{A}\nabla\phi^h) \cdot \mathbb{A}^{-1}(\widehat{\mathbf{u}}^h + \mathbb{A}\nabla\widehat{\phi}^h) \, d\Omega + \int_\Omega \gamma^{-1}(\nabla \cdot \mathbf{u}^h + \gamma\phi^h)(\nabla \cdot \widehat{\mathbf{u}}^h + \gamma\widehat{\phi}^h) \, d\Omega \, , \end{aligned} \tag{9}$$

$$P(\{\widehat{\phi}^h, \widehat{\mathbf{u}}^h\}) = \int_\Omega \gamma^{-1} f(\nabla \cdot \widehat{\mathbf{u}}^h + \gamma\widehat{\phi}^h) \, d\Omega \, . \tag{10}$$

To show that Q is coercive on $\mathbf{G}^h \times \mathbf{D}^h$, note that

$$\begin{aligned}
&Q(\{\phi^h, \mathbf{u}^h\}, \{\widehat{\phi}^h, \widehat{\mathbf{u}}^h\}) \\
&\quad = \int_\Omega \gamma^{-1} \nabla \cdot \mathbf{u}^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \mathbf{u}^h \cdot \mathbb{A}^{-1} \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \nabla \phi^h \cdot \mathbb{A} \nabla \widehat{\phi}^h \, d\Omega + \int_\Omega \gamma \phi \phi^h \, d\Omega \\
&\qquad + \int_\Omega \mathbf{u}^h \cdot \nabla \widehat{\phi}^h \, d\Omega + \int_\Omega \nabla \cdot \mathbf{u}^h \widehat{\phi}^h \, d\Omega + \int_\Omega \nabla \phi^h \cdot \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \phi^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega \\
&\quad = \int_\Omega \gamma^{-1} \nabla \cdot \mathbf{u}^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \mathbf{u}^h \cdot \mathbb{A}^{-1} \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \nabla \phi^h \cdot \mathbb{A} \nabla \widehat{\phi}^h \, d\Omega + \int_\Omega \gamma \phi \phi^h \, d\Omega \, .
\end{aligned}$$

From the last identity, (2), and $\gamma \geq \gamma_0 > 0$, it is clear that there is a constant C such that

$$\begin{aligned}
&Q(\{\phi^h, \mathbf{u}^h\}, \{\phi^h, \mathbf{u}^h\}) = \\
&\quad \|\mathbb{A}^{-1/2} \mathbf{u}^h\|_0^2 + \|\gamma^{-1/2} \nabla \cdot \mathbf{u}^h\|_0^2 + \|\gamma^{1/2} \phi^h\|_0^2 + \|\mathbb{A}^{1/2} \nabla \phi^h\|_0^2 \geq C \Big(\|\phi^h\|_1^2 + \|\mathbf{u}^h\|_{H(\Omega, \mathrm{div})}^2 \Big) \, .
\end{aligned}$$

Using the same identity, it is easy to see that (8) *decouples* into the two independent equations

$$\int_\Omega \gamma^{-1} \nabla \cdot \mathbf{u}^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \mathbf{u}^h \mathbb{A}^{-1} \widehat{\mathbf{u}}^h \, d\Omega = \int_\Omega \gamma^{-1} f \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega \quad \forall \widehat{\mathbf{u}}^h \in \mathbf{D}^h \tag{11}$$

$$\int_\Omega \nabla \phi^h \mathbb{A} \nabla \widehat{\phi}^h \, d\Omega + \int_\Omega \gamma \phi^h \widehat{\phi}^h \, d\Omega = \int_\Omega f \widehat{\phi}^h \, d\Omega \qquad \forall \widehat{\phi}^h \in \mathbf{G}^h \tag{12}$$

for $\mathbf{u}^h$ and ϕ^h, respectively. The scalar variable ϕ^h is completely determined by the second equation, which is identical to (3). To show that $\mathbf{u}^h$ is the same as in (4), note that if $\gamma \geq \gamma_0$, the second equation in (4) can be restated in the form

$$\int_\Omega \lambda^h \widehat{\lambda}^h \, d\Omega = \int_\Omega \frac{1}{\gamma} f \widehat{\lambda}^h \, d\Omega - \int_\Omega \frac{1}{\gamma} \nabla \cdot \mathbf{u}^h \widehat{\lambda}^h \, d\Omega \quad \forall \widehat{\lambda}^h \in \mathbf{Q}^h \, .$$

The fact that $\mathbf{Q}^h = \nabla \cdot (\mathbf{D}^h)$ implies that for any $\widehat{\lambda}^h \in \mathbf{Q}^h$ there exists $\widehat{\mathbf{u}}^h \in \mathbf{D}^h$ such that $\widehat{\lambda}^h = \nabla \cdot \widehat{\mathbf{u}}^h$ and so the last equation becomes

$$\int_\Omega \lambda^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega = \int_\Omega \frac{1}{\gamma} f \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega - \int_\Omega \frac{1}{\gamma} \nabla \cdot \mathbf{u}^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega \quad \forall \widehat{\mathbf{u}}^h \in \mathbf{D}^h \, .$$

This enables us to eliminate λ^h from the first equation in (4) and to obtain a formulation in terms of $\mathbf{u}^h$ only: seek $\mathbf{u}^h \in \mathbf{D}^h$ such that

$$\int_\Omega \gamma^{-1} \nabla \cdot \mathbf{u}^h \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega + \int_\Omega \mathbb{A}^{-1} \mathbf{u}^h \cdot \widehat{\mathbf{u}}^h \, d\Omega = \int_\Omega \gamma^{-1} f \nabla \cdot \widehat{\mathbf{u}}^h \, d\Omega \quad \forall \widehat{\mathbf{u}}^h \in \mathbf{D}^h \, . \tag{13}$$

This last equation is identical to (11).

To prove the local conservation property of $\mathbf{u}^h$ we reverse the elimination process used to obtain (13) by solving the equation: seek $\phi^h \in \mathbf{Q}^h$ such that

$$\int_\Omega \lambda^h \widehat{\lambda}^h \, d\Omega = \int_\Omega \gamma^{-1} f \widehat{\lambda}^h \, d\Omega - \int_\Omega \gamma^{-1} \nabla \cdot \mathbf{u}^h \widehat{\lambda}^h \, d\Omega \quad \forall \widehat{\lambda}^h \in \mathbf{Q}^h \, . \tag{14}$$

to determine λ^h. Because $\mathbf{Q}^h$ is discontinuous, this problem breaks down into local element problems which are easy to solve. To show that λ^h and $\mathbf{u}^h$ verify (7), note that (14) can be written as

$$\int_\Omega \Big(\nabla \cdot \mathbf{u}^h + \gamma \lambda^h - f \Big) \gamma^{-1} \widehat{\lambda}^h \, d\Omega = 0 \quad \forall \widehat{\lambda}^h \in \mathbf{Q}^h$$

which is the definition of $\Pi_{\mathbf{Q}^h}$. This also proves that $(\mathbf{u}^h, \lambda^h)$ solves the mixed problem (4). □

Several comments about (5) are in order. First we stress again that $\mathbf{G}^h$ and $\mathbf{D}^h$ are chosen to correspond to spaces used for ϕ^h and $\mathbf{u}^h$ in the Ritz-Galerkin and the mixed Galerkin methods, respectively, but that they are in no way required to be connected to each other. A notable and very attractive computational property of (5) is that, effectively, it simultaneously solves the Ritz-Galerkin (3) and the mixed-Galerkin (4) problems. Furthermore, the computation of ϕ^h and $\mathbf{u}^h$ requires solution of a symmetric and positive definite linear system which is much easier to do than to solve the indefinite system arising from (4).

For a comparison, a least-squares implementation that uses equal order C^0 finite elements is associated with the least-squares principle

$$\min_{\phi^h \in \mathbf{G}^h, \mathbf{u}^h \in (\mathbf{G}^h)^n} J(\phi^h, \mathbf{u}^h; f)\,. \tag{15}$$

In (15) the choice of finite elements for ϕ^h is compatible with (3) but the choice of finite elements for $\mathbf{u}^h$ is not compatible with (4). Section 4 contains examples that illustrate suboptimal convergence rates of (15) and the problems arising in approximation of non-smooth solutions.

3. Compatible least-squares finite element methods for $\gamma = 0$

When $\gamma = 0$ the compatible least-squares functional (6) is not defined. To handle this case we introduce the modified least-squares functional

$$J_0(\phi^h, \mathbf{u}^h; f) = \frac{1}{2}\left(\|\mathbb{A}^{-1/2}(\mathbf{u}^h + \mathbb{A}\nabla\phi^h)\|_0^2 + \|\nabla\cdot\mathbf{u}^h - f\|_0^2\right)$$

and the minimization problem

$$\min_{\phi^h \in \mathbf{G}^h, \mathbf{u}^h \in \mathbf{D}^h} J_0(\phi^h, \mathbf{u}^h; f)\,. \tag{16}$$

This problem retains most of the attractive computational properties of (5), including the optimal error estimates for both variables [2]. However, the proof of local conservation in Theorem 1 does not carry over to the case $\gamma = 0$ because the weak variational equation for (16) does not decouple into two independent equations for ϕ^h and $\mathbf{u}^h$, respectively. It turns out that local conservation can be easily recovered by a simple flux-correction procedure.

3.0.1. Flux correction When $\gamma = 0$, solution of (16) gives $\mathbf{u}^h \in \mathbf{D}^h$ such that

$$\int_\Omega \nabla\cdot\mathbf{u}^h\widehat{\phi}^h\,d\Omega \approx \int_\Omega f\widehat{\phi}^h\,d\Omega \quad \forall\widehat{\phi}^h \in \mathbf{Q}^h\,.$$

In what follows we formulate a simple, *local* procedure that replaces $\mathbf{u}^h$ by a field $\widetilde{\mathbf{u}}^h \in \mathbf{D}^h$ that satisfies the above equation exactly. In particular, if $f = 0$, the field $\widetilde{\mathbf{u}}^h$ will be divergence free at any point $\mathbf{x}$ in the interior of an element K.

For the sake of clarity, we describe the procedure for $f = 0$, quadrilateral elements, and the lowest-order elements $\mathbf{D}^h$ and $\mathbf{Q}^h$; see [5]. Let K denote an arbitrary quadrilateral element. A

function $\mathbf{u}^h \in \mathbf{D}^h$ on this element can be written in the form

$$\mathbf{u}^h = \sum_{i=1}^{4} \Phi_{\mathcal{F}_i} W_{\mathcal{F}_i} ,$$

where the $\mathcal{F}_i$'s are the (oriented) faces of K, $\Phi_{\mathcal{F}_i}$ is the flux of $\mathbf{u}^h$ across the face $\mathcal{F}_i$, and $W_{\mathcal{F}_i}$ is the shape function associated with face $\mathcal{F}_i$. We recall that (see, e.g., [3])

$$\nabla \cdot \mathbf{u}^h{}_{|K} = (\sigma_1 \Phi_{\mathcal{F}_1} + \sigma_2 \Phi_{\mathcal{F}_2} + \sigma_3 \Phi_{\mathcal{F}_3} + \sigma_4 \Phi_{\mathcal{F}_4}) W_K ,$$

where W_K is the (single) basis function of $\mathbf{Q}^h$ on K and $\sigma_i = 1$ if orientation of $\mathcal{F}_i$ coincides with the outer normal on ∂K and $\sigma_i = -1$ otherwise. We seek to define modified flux values $\widetilde{\Phi}_{\mathcal{F}_i} = \Phi_{\mathcal{F}_i} - \sigma_i \Delta\Phi_{\mathcal{F}_i}$ and a function

$$\widetilde{\mathbf{u}}^h = \sum_{i=1}^{4} \widetilde{\Phi}_{\mathcal{F}_i} W_{\mathcal{F}_i}$$

such that

$$\sigma_1 \widetilde{\Phi}_{\mathcal{F}_1} + \sigma_2 \widetilde{\Phi}_{\mathcal{F}_2} + \sigma_3 \widetilde{\Phi}_{\mathcal{F}_3} + \sigma_4 \widetilde{\Phi}_{\mathcal{F}_4} = 0 .$$

To define the corrected flux values for an element K, we proceed as follows. If, for a given face $\mathcal{F}_i$, the flux $\Phi_{\mathcal{F}_i}$ has already been corrected, we set $\Delta\Phi_{\mathcal{F}_i} = 0$. If the flux on $\mathcal{F}_i$ has not yet been corrected, we set

$$\Delta\Phi_{\mathcal{F}_i} = \frac{\sigma_1 \Phi_{\mathcal{F}_1} + \sigma_2 \Phi_{\mathcal{F}_2} + \sigma_3 \Phi_{\mathcal{F}_3} + \sigma_4 \Phi_{\mathcal{F}_4}}{n(K)} ,$$

where $n(K) > 0$ is the total number of faces on K whose fluxes have not been corrected.

Consider now a partition $\mathcal{T}_h$ of Ω with n_h elements. To define the flux-correction algorithm, some additional notation is necessary. Let

$$\mathcal{F}(i_1, \ldots, i_k) = \{\mathcal{F} \,|\, \mathcal{F} \in \partial K_i, i = i_1, \ldots, i_k\}$$

denote the set of all faces in the union of the elements indexed by $i_1, \ldots, i_k$. For example, $\mathcal{F}(i_l) = \partial K_{i_l}$ is the set of all faces of element K_{i_l}. Define

$$n(i_1, \ldots, i_k) = \dim\Big\{\mathcal{F}(i_k)/\{\mathcal{F}(i_1, \ldots, i_{k-1}) \cap \mathcal{F}(i_k)\}\Big\} \geq 0 .$$

Given $\mathbf{u}^h \in \mathbf{D}^h$, the following algorithm finds $\widetilde{\mathbf{u}}^h \in \mathbf{D}^h$ that is divergence free:

1. Define a permutation $\pi = \{i_1, i_2, \ldots, i_{nel}\}$ of all elements in $\mathcal{T}_h$, such that

$$n(i_1, \ldots, i_k) > 0 \quad \text{for } k = 1, \ldots, n_h ;$$

2. for $k = 1, \ldots, n_h$, apply the element flux correction procedure to element K_{i_k}.

3.0.2. Error estimates For the sake of completeness, we provide, in the following theorem proved in [2], optimal error estimates. Recall that $\mathbf{G}^h$ and $\mathbf{D}^h$ are in no way connected to each other and can be selected independently. In fact, as has already been pointed out, they can even be defined with respect to separate triangulations. We assume that $\mathbf{G}^h$ is one of the

standard nodal C^0 spaces with the property that there exists an integer $m > 0$ such that for every $\phi \in H^{m+1}(\Omega)$ there is a function $\phi^h \in \mathbf{G}^h$ such that

$$\|\phi - \phi^h\|_0 + h\|\nabla\phi - \nabla\phi^h\|_0 \le Ch^{m+1}\|\phi\|_{m+1}\,. \tag{17}$$

Regarding $\mathbf{D}^h$ we assume that it is from one of the RT_{k-1} or BDM_k; $k > 0$ families of finite elements. We recall that for any $\mathbf{u} \in H(\Omega, \mathrm{div}\,) \cap \mathbf{H}^{k+1}(\Omega)$ there exists a $\mathbf{u}^h \in \mathbf{D}^h$ such that

$$\|\mathbf{u} - \mathbf{u}_h\|_0 \le C \begin{cases} h^k\|\mathbf{u}\|_k & \text{if } \mathbf{D}^h = \mathrm{RT}_{k-1} \\ h^{k+1}\|\mathbf{u}\|_{k+1} & \text{if } \mathbf{D}^h = \mathrm{BDM}_k\,. \end{cases} \quad ; \quad \|\nabla\cdot(\mathbf{u} - \mathbf{u}^h)\|_0 \le Ch^k\|\mathbf{u}\|_{k+1}\,. \tag{18}$$

Theorem 2. *Let $(\phi^h, \mathbf{u}^h) \in \mathbf{G}^h \times \mathbf{D}^h$ solve the minimization problem*

$$\min_{\phi^h\in\mathbf{G}^h, \mathbf{u}^h\in\mathbf{D}^h} J_C(\phi^h, \mathbf{u}^h; f)\,, \quad \textit{where} \quad J_C(\phi^h, \mathbf{u}^h; f) = \begin{cases} J(\phi^h, \mathbf{u}^h; f) & \textit{if } \gamma \ge \gamma_0 > 0 \\ J_0(\phi^h, \mathbf{u}^h; f) & \textit{if } \gamma = 0 \end{cases} \tag{19}$$

Assume that (1) has full elliptic regularity and that $(\phi, \mathbf{u}) \in H^1_D(\Omega) \cap H^{r+1}(\Omega) \times H_N(\Omega, \mathrm{div}\,) \cap \mathbf{H}^{r+1}(\Omega)$ for some integer $r \ge 1$. Also assume that the approximation orders of $\mathbf{G}^h$ and $\mathbf{D}^h$ are equilibrated with respect to the solution regularity, i.e., $m = k = r$. Then, there exists a constant $C > 0$ such that

$$\|\phi - \phi_h\|_0 + \|\nabla\phi - \nabla\phi^h\|_0 \le Ch^{r+1}(\|\mathbf{u}\|_{r+1} + \|\phi\|_{r+1})\,, \tag{20}$$

$$\|\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}^h\|_0 \le Ch^{r}(\|\mathbf{u}\|_{r+1} + \|\phi\|_{r+1})\,, \tag{21}$$

and

$$\|\mathbf{u} - \mathbf{u}_h\|_0 \le C \begin{cases} h^r\Big(\|\mathbf{u}\|_{r+1} + \|\phi\|_{r+1}\Big) & \textit{if } \mathbf{D}^h = \mathrm{RT}_{r-1} \\ h^{r+1}\Big(\|\mathbf{u}\|_{r+1} + \|\phi\|_{r+1}\Big) & \textit{if } \mathbf{D}^h = \mathrm{BDM}_r\,. \end{cases} \tag{22}$$

4. Computational studies

In all experiments Ω is the unit square in $\mathbb{R}^2$. Finite element spaces are defined on logically Cartesian (but not necessarily rectangular) partitions of Ω into quadrilaterals. Implementation of the methods requires the spaces $(\mathbf{G}^h, \mathbf{D}^h)$ for the compatible least-squares formulation (19), the spaces $(\mathbf{D}^h, \mathbf{Q}^h)$ for the mixed Galerkin method (4), and the space $\mathbf{G}^h$ for the Ritz-Galerkin method (3). In all but one example we use the spaces $\mathbf{G}^h = Q_1$, $\mathbf{D}^h = RT_0$, and $\mathbf{Q}^h = Q_0$. For these spaces (20)-(22) specialize to

$$\|\phi - \phi^h\|_0 + h\|\nabla\phi - \nabla\phi^h\|_0 \le Ch^2(\|\phi\|_2 + \|\mathbf{u}\|_2)$$

and

$$\|\mathbf{u} - \mathbf{u}^h\|_0 + \|\nabla\cdot\mathbf{u} - \nabla\cdot\mathbf{u}^h\|_0 \le Ch(\|\phi\|_2 + \|\mathbf{u}\|_2)\,,$$

respectively. The matrices in all methods are assembled using a 2×2 Gauss quadrature rule. The linear systems are solved "exactly" using direct solvers from LAPACK.

Table I. Convergence rates of the nodal $\mathbf{G}^h \times (\mathbf{G}^h)^2$ implementation vs. best approximation rates.

Variable	scalar		vector	
Error	L^2	H^1	L^2	$H(\Omega, \mathrm{div})$
Q1	2.00	1.00	**1.38**	0.99
BA	2.00	1.00	2.00	1.00
Q2	3.00	2.00	**2.02**	2.00
BA	3.00	2.00	3.00	2.00

4.1. Convergence of the nodal least-squares method

Convergence rates of (15) are estimated for (1) with $\mathbb{A} = \mathbb{I}$ and $\gamma = 0$ by computing approximations of the manufactured solution $\phi = -e^x \sin y$ and $\mathbf{u} = -\nabla\phi$ on a sequence of uniform grids. For this solution $\nabla \cdot \mathbf{u} = 0$. Table I shows the estimated convergence rates of (15), implemented using the spaces $\mathbf{G}^h = Q_1$ and $\mathbf{G}^h = Q_2$. We see that convergence rates of the pressure in L^2 norm and the H^1 seminorm coincide with the best theoretical approximation rates of the finite element spaces. Convergence of the velocity approximation is optimal with respect to the $H(\Omega, \mathrm{div})$ norm. However, the L^2 convergence of that variable is clearly suboptimal. The optimal $H(\Omega, \mathrm{div})$ rates are due to the equivalence of (6) and the inner product on $H^1_D(\Omega) \times H_N(\Omega, \mathrm{div})$, which remains valid on the proper subspace $\mathbf{G}^h \times (\mathbf{G}^h)^2$.

4.2. Nodal vs. compatible least squares for non-smooth solutions

The following example from [10] is used to demonstrate the troubles in the nodal least-squares method (15) arising from non-smooth solutions of (1) and to underscore the excellent performance of the compatible method. In this example Ω is divided into five identical horizontal strips $\Omega_i = \{(x, y) \,|\, 0.2(i-1) \le y < 0.2i;\ 0 \le x \le 1\}$ and $\mathbb{A} = \kappa\mathbb{I}$ where κ is a piecewise constant function taking the value κ_i on strip Ω_i. We take $\gamma = 1$, $\kappa_1 = 16$, $\kappa_2 = 6$, $\kappa_3 = 1$, $\kappa_4 = 10$, $\kappa_5 = 2$, and set the right hand side so that the exact solution of (1) is given by $\phi = 1 - x$ and $\mathbf{u} = (\kappa_i, 0)$ on strip Ω_i. Therefore, the normal component of $\mathbf{u}$ is continuous across the interface between any two of the strips, but its tangential component is discontinuous along this interface.

Table II. Errors in the compatible and nodal least-squares solutions

Variable	scalar		vector	
Error	L^2	H^1	L^2	$H(\Omega, \mathrm{div})$
Compatible	0.1847E-12	0.2460E-13	0.1789E-10	0.1847E-12
Nodal	0.8892E-02	0.1423E+00	0.1925E+01	0.7206E+02

Figure 1 shows velocity and pressure contours computed by the compatible and the nodal

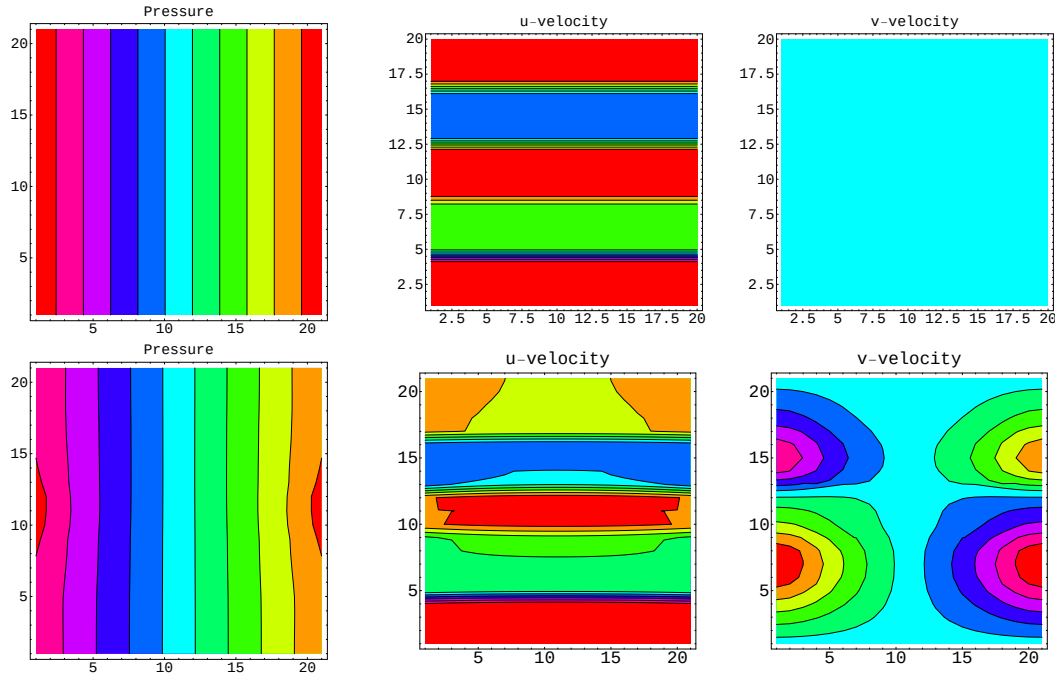

Figure 1. Pressure and velocity contours computed by the compatible least-squares method (top row) and by the nodal least-squares method (bottom row).

least-squares methods. From the plots we see that the compatible method has recovered the exact solution. In contrast, the nodal method shows large qualitative difference in the velocity components. It also fails to recover the pressure despite the fact that it belongs to the bilinear finite element space Q_1. Figure 2 shows surface plots of the velocity field components and further highlights the substantial error in the nodal velocity approximation. The conclusions drawn from Figs. 1-2 are asserted by the error data in Table II.

4.3. Comparison with Ritz-Galerkin and mixed Galerkin methods

The example in this section illustrates the conclusions of Theorem 1. We consider (1) with $\mathbb{A} = \mathbb{I}$, $\gamma = \gamma_0 > 0$, and the manufactured solution $\phi_0 = -\sin(\pi x^{1.25})^2 \sin(\pi y^{1.5})^2$, $\mathbf{u}_0 = \nabla\phi_0$. This solution satisfies homogeneous Dirichlet and Neumann boundary conditions. Table III compares the errors of the compatible least-squares method with the errors of the Ritz-Galerkin (3) and the mixed-Galerkin (4) finite element methods on a sequence of uniform grids. From this table we see that the velocity errors of the compatible least-squares and the mixed method are identical, i.e., the two methods compute the same velocity field. Likewise, the pressure errors of the compatible least-squares method and the Ritz-Galerkin method are the same indicating that the two methods have computed identical approximations.

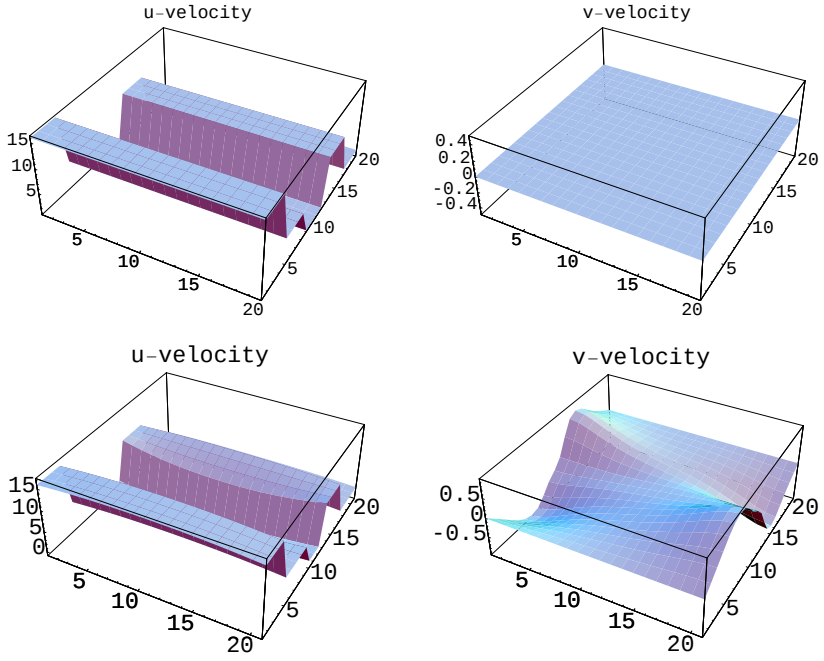

Figure 2. Velocity surface plots of the compatible least-squares method (top row) and the nodal least-squares method (bottom row).

Table III. Compatible least-squares vs. Ritz-Galerkin (3) and mixed Galerkin (4) methods for $\gamma > 0$.

error	method	16	32	64	128
L^2	LS	0.1514803E+00	0.7192623E-01	0.3523105E-01	0.1745720E-01
vector	KP	0.1514803E+00	0.7192623E-01	0.3523105E-01	0.1745720E-01
$H(\Omega, \mathrm{div})$	LS	0.2869324E+01	0.1397179E+01	0.6894290E+00	0.3426716E+00
vector	KP	0.2869324E+01	0.1397179E+01	0.6894290E+00	0.3426716E+00
L^2 scalar	DP	0.3997943E-02	0.9378368E-03	0.2274961E-03	0.5621838E-04
	LS	0.3997943E-02	0.9378368E-03	0.2274961E-03	0.5621838E-04
	KP	0.3679584E-01	0.1778803E-01	0.8750616E-02	0.4340574E-02
$H^1(\Omega)$	DP	0.2671283E+00	0.1296329E+00	0.6383042E-01	0.3166902E-01
scalar	LS	0.2671283E+00	0.1296329E+00	0.6383042E-01	0.3166902E-01

4.4. Flux correction procedure

To illustrate the flux correction algorithm from Section 3.0.1, we consider (1) with $\mathbb{A} = \mathbb{I}$, $\gamma = 0$ and the first manufactured solution for which $\mathbf{u} = -\nabla\phi$ was solenoidal. Table IV compares

the L^2 and divergence errors of the velocity computed by (16) before and after application of the flux correction (rows "LS" and "LS-FC", respectively), with the errors of the velocity field computed by the mixed method (4) (row "Mixed"). Errors are reported for uniform and randomly perturbed grids with 900 elements. The first two rows in Table IV show that

Table IV. Performance of the flux correction procedure.

Grid	30×30 Uniform grid		30×30 Random grid	
Error	L^2 error	$\|\nabla \cdot \mathbf{u}^h\|_0$	L^2 error	$\|\nabla \cdot \mathbf{u}^h\|_0$
LS	0.2145816E-01	0.1235067E-03	0.2342838E-01	0.1668777E-03
LS-FC	0.2145813E-01	0.4612541E-14	0.2342836E-01	0.5281635E-14
Mixed	0.2145805E-01	0.1446273E-13	0.2352581E-01	0.6658244E-14

postprocessing can recover local conservation without having any effect on the L^2 accuracy of the vector field. The divergence of $\mathbf{u}^h$ drops to machine precision, while the L^2 errors before and after the flux correction are virtually identical.

The last two rows in Table IV show that, with regard to accuracy and local conservation, (16) with flux correction is identical in performance to the mixed-Galerkin method. The L^2 errors of the vector field approximations obtained by the two methods are the same to five digits on the uniform mesh and to three digits on the random mesh. The divergence of the postprocessed vector approximation is actually closer to machine precision than that of the mixed method solution. This conclusion is further upheld by an examination of both methods on a sequence of uniform grids. Table V contains error values for selected grid sizes. The data show not only identical convergence rates for the vector variable approximations in both methods, but also asymptotically identical values of the errors for that variable.

Table V. Compatible least-squares with flux-corrected velocity vs. mixed-Galerkin method.

error	method	16	32	64	128
L^2	LS-FC	4.148515E-02	2.007374E-02	9.877587E-03	4.899909E-03
vector	Mixed	4.148467E-02	2.007368E-02	9.877579E-03	4.899908E-03
$H(\Omega, \mathrm{div})$	LS-FC	2.491389E-15	5.181631E-15	1.032372E-14	2.124581E-14
vector	Mixed	4.664771E-15	2.738185E-14	1.817327E-14	4.383082E-14
L^2	LS-FC	3.607090E-04	8.436735E-05	2.042191E-05	5.025008E-06
scalar	Mixed	3.439163E-02	1.664313E-02	8.189706E-03	4.062637E-03

5. Conclusions

A compatible least-squares finite element method that inherits the best computational properties of the Ritz-Galerkin and the mixed Galerkin methods for (1) has been formulated and studied numerically. When $\gamma > 0$ the compatible least-squares method provides a *simultaneous* solution of the Ritz-Galerkin and the mixed-Galerkin methods and its solution is locally conservative. When $\gamma = 0$, (19) yields velocity approximations that are very close to those in the mixed method. An additional flux-correction step can be applied to make the vector approximation locally conservative without compromising its L^2 accuracy.

In addition to these attractive features, (19) retains the valuable property of leading to symmetric and positive definite linear systems. Of course, one can consider the nodal implementation (15) in lieu of (19). This choice offers a somewhat simpler implementation, symmetric positive definite linear systems and is compatible with respect to the origins of the scalar variable. As a result, it leads to optimally accurate approximations for that variable. However, considering that the principal reason to introduce the vector variable is to obtain more accurate approximations, it is clear that the nodal least-squares method (15) falls short of this task. Not only does the nodal vector approximation converge at suboptimal L^2 rates, but it is also fails to be locally conservative regardless of the value of γ. Moreover, its lack of local conservation cannot be remedied in any way.

REFERENCES

1. P. Bochev and M. Gunzburger, Least-squares finite element methods for elliptic equations, *SIAM Review*, 40, pp. 789–837, 1998.
2. P. Bochev and M. Gunzburger, On least-squares finite elements for the Poisson equation and their connection to the Kelvin and Dirichlet principles, *SIAM J. Num. Anal.* 43/1. pp. 340-362, 2005.
3. P. Bochev and A. Robinson; Matching algorithms with physics: exact sequences of finite element spaces; in: *Collected Lectures on the Preservation of Stability Under Discretization,* edited by D. Estep and S. Tavener, SIAM, Philadelphia, 2001.
4. F. Brezzi, On existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers, *RAIRO Model. Math. Anal. Numer.*, 21, pp. 129–151, 1974.
5. F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer, Berlin, 1991.
6. Z. Cai, T. Manteuffel, and S. McCormick, First-order system least-squares for second order partial differential equations: Part II, *SIAM J. Numer. Anal.*, 34, pp. 425-454, 1997.
7. G. Carey and A. Pehlivanov Error estimates for least-squares mixed finite elements, *Math. Mod. Numer. Anal.*, 28, pp. 499–516, 1994.
8. C. Chang, Finite element approximation for grad-div type systems in the plane, *SIAM J. Numer. Anal.*, 29, pp. 452–461, 1992.
9. G. Fix, M. Gunzburger and R. Nicolaides; On finite element methods of the least-squares type; *Comput. Math. Appl.* 5, pp. 87-98, 1979.
10. Hughes, T.J.R., A. Masud and J. Wan; A Stabilized Mixed Discontinuous Galerkin Method for Darcy Flow, in press, *Comp. Meth. Appl. Mech. Engrg*, special issue on discontinuous Galerkin methods.
11. D. Jesperson; A least-squares decomposition method for solving elliptic equations, *Math. Comp.*, 31, 873-880, 1977.
12. B. N. Jiang and L. Povinelli, *Optimal least-squares finite element methods for elliptic problems,* Comp. Meth. Appl. Mech. Engrg., 102, pp. 199–212, 1993.